

## El Modelo de Respuesta no Aleatorizada Transversal: Una Aplicación

*Olga Solano*<sup>1</sup>

*Jhon Galván*<sup>2</sup>

(Recibido: 28/03/2016 - Aceptado: 21/04/2016)

**Resumen:** En el presente trabajo presentamos los resultados de una investigación realizada en la Facultad de Ciencias Matemáticas de la UNMSM, para investigar la proporción de personas que han copiado en los exámenes por lo menos una vez, el porcentaje de personas que han utilizado notas en un examen, el porcentaje de personas que han utilizado drogas para mejorar su rendimiento académico y el porcentaje de personas que han presentado un documento que contienen un párrafo intencionalmente adoptado del trabajo de otra persona, utilizando la teoría del modelo de respuesta no aleatorizada transversal (Tian & Tang 2014). En el diseño muestral, se utilizó el muestreo aleatorio estratificado con afijación proporcional al tamaño de cada estrato (Scheffer & Mendenhall, 2007) considerando como estratos a las Escuelas Académicas Profesionales (E.A.P.) de la FCM. Para el cálculo del tamaño de muestra se consideró un límite para el error de estimación de 5%, con un nivel de confianza del 95% y la información proporcionada por la Dirección Académica de la FCM, de los alumnos matriculados el primer semestre del año académico 2015, el tamaño de muestra fue de 300 alumnos, repartidos en forma proporcional a las cuatro E.A.P. de la FCM.

**Palabras Claves:** Modelo de respuesta no aleatorizada transversal, tamaño de muestra, pregunta delicada, pregunta no relacionada.

### Response Model non Randomized Cross: An Application

**Abstract:** In this paper we present the results of research conducted in the Faculty of Mathematics (FCM) at the University Nacional Mayor de San Marcos in Peruvian to investigate the proportion of people who have copied examinations at last once, percentage of people who have used notes in an exam, percentage of people who have used drugs to improve their academic performance and percentage of people who have presented a document containing an intentionally adopted the work of another paragraph, using the theory the nonrandomized response model, the cross model (Tian & Tang 2014). Stratified random sampling with allocation proportional to the size of each stratum was used in the sample design (Scheffer & Mendenhall, 2007), considering as strata to the Professional Academic School (E.A.P) of the FCM. For the calculation of sample size limit to the estimation error of 5%, with a confidence level of 95% and the information provided by the Academic Board of the FCM, of the students enrolled in the first half of the year it was considered academic 2015, the sample size was 300 students, distributed in proportion to the four EAP of the FCM.

**Key Words:** Response model nonrandomized, cross model, delicate question, unrelated question.

---

<sup>1</sup>UNMSM, Facultad de Ciencias Matemáticas, e-mail:osolanod@unmsm.edu.pe

<sup>2</sup>UNMSM, Facultad de Ciencias Matemática, e-mail: jgalvan.tapia@gmail.com

## 1 Introducción

Existen multitud de estudios realizados para mejorar la calidad y veracidad de las respuestas obtenidas sobre temas sensibles, como consumo de drogas, relaciones sexuales con más de dos personas, consumo de alcohol, copia de exámenes, etc. Frecuentemente los investigadores obtienen resistencias a la participación en estos estudios u obtienen respuestas falsas de parte de los entrevistados. Warner (1965) realizó la primera propuesta para obtener respuestas válidas ante preguntas embarazosas basándose en la realización de dos preguntas mutuamente excluyentes (por ejemplo “A: Declaré mis ingresos extraordinarios el año pasado”; “B: No declaré mis ingresos extraordinarios el año pasado”). Posteriormente, se desarrollaron otros métodos basados en Warner, 1965, como el denominado “método de alternativa forzada” de Tracy (Fox & Tracy, 1986). El modelo de respuesta aleatorizada (MRNA) es un método especialmente diseñado para asegurar privacidad a los entrevistados en el estudio de temas sensibles, delicados o embarazosos. Se intenta con ello evitar el sesgo de los entrevistados en ciertas conductas hacia la respuesta socialmente más deseable. Se ha utilizado para analizar temas como copiar en los exámenes, insolvencia, fraudes, haber sido arrestado, conducir bajo los efectos del alcohol, tener un hijo fuera del matrimonio, aborto, etc.

Yu, Tian y Tang (2008) propuso el modelo de respuesta no aleatorizada llamado el modelo trasnversal, para estimar la proporción desconocida  $\pi = P(Y = 1)$ , de personas con una característica sensible, mediante la introducción de una variable aleatoria dicotómica no sensible  $W$  tal que  $Y$  y  $W$  son mutuamente independientes, Yu, Tian y Tang(2008) ha desarrollado un marco general de análisis y diseño para el modelo transversal de respuesta no aleatorizada.

El MRNA, el cual utiliza una o dos variables no sensibles (ejemplo, fecha de nacimiento del entrevistado o el último dígito del número de teléfono del entrevistado), combinado con una o dos variables sensibles para formar una tabla de contingencia incompleta y obtener indirectamente de los entrevistados respuestas sensibles (Takahasi & Sakasegawa, 1977); (Tian, Yu, Tang, & Geng, 2007b); (Tian, Tang, Liu, Tan, & Tang, 2011); (Yu, Tian,& Tang, 2008); (Tan, Tian, & Tang, 2009); (Tang, Tian, Tang, & Liu, 2009), el diseño no requiere de mecanismos de aleatorización. El objetivo del presente trabajo es presentar los resultados de una investigación realizada en la FCM de la UNMSM utilizando el modelo de respuesta no aleatorizada, el modelo transversal, en el contexto descrito con la finalidad de que cualquier investigador interesado en usar esta metodología conozca los aspectos fundamentales de la misma.

## 2 Metodología

### El modelo de respuesta no aleatorizada - el modelo transversal

Yu, Tian y Tang (2008) hizo una propuesta del modelo transversal, el cual es una versión del modelo de la pregunta no relacionada. Sea  $\{Y = 1\}$  denota una clase de personas que tienen una característica (por ejemplo, consumo de drogas) y  $\{Y = 0\}$  denota la clase complementaria. El objetivo es estimar la proporción de los elementos que pertenecen a la clase  $\{Y = 1\}$ ,  $\pi = P(Y = 1)$ . Suponga que  $W$  es una variable dicotómica no sensible,  $Y$  y  $W$  son independientes entre sí con  $p = P(W = 1)$  conocida.

### El diseño de la encuesta para el modelo trasnversal

Los entrevistadores pueden diseñar un cuestionario en el formato mostrado en el lado izquierdo de la Tabla 1, y pedir a cada entrevistado que marque con una cruz en el círculo superior, si él/ella pertenece a uno de los dos círculos o marque con una cruz en el cuadrado superior si él/ella pertenece a uno de los dos cuadrados, por lo que el entrevistado no está siendo expuesto si la cruz

esta puesta en la parte superior del círculo o cuadrado, note que todas las clases  $\{Y = 0, W = 0\}$ ,  $\{Y = 0, W = 1\}$  son clases no sensibles. Así, un entrevistado que pertenece a la clase sensible (es decir,  $\{Y = 1\}$ ) no está siendo expuesto si marca en el círculo/cuadrado superior. Por lo tanto, si él entrevistado pertenece a la clase sensible no es conocido por los entrevistados. Las probabilidades de las celdas correspondientes son mostradas en el lado derecho de la Tabla 1. Las dos variables binarias  $Y$  y  $W$  son independientes, la probabilidad conjunta es el producto de dos probabilidades marginales correspondientes.

Cuadro 1: El modelo transversal y las correspondientes celdas de probabilidades

Categoría	W=0	W=1	Categoría	W=0	W=1	Marginal
Y=0	○	□	Y=0	$(1 - \pi)(1 - p)$	$(1-\pi)p$	$1-\pi$
Y=1	□	○	Y=1	$\pi(1 - p)$	$\pi p$	$\pi$
			Marginal	$1-p$	$p$	$1$

En relación a la Tabla 1, se define una variable aleatoria de Bernoulli como

$$Y^T = \begin{cases} 1, & \text{si la marca es en el círculo superior} \\ 0, & \text{si la marca es en el cuadrado superior} \end{cases}$$

Donde el superíndice “ T ” representa a la variable de Bernoulli para el modelo transversal. Por lo tanto, las probabilidades de  $Y^T = 1$  y  $Y^T = 0$  están dadas por:  $P[Y^T = 1] = (1 - \pi)(1 - p) + \pi p$  y  $P[Y^T = 0] = \pi(1 - p) + (1 - \pi)p$ , respectivamente.

Sea  $Y_{obs} = \{y_i^T : i = 1, \dots, n\}$  denota los datos observados para los  $n$  encuestados, entonces la función de probabilidad para  $\pi$  es:

$$L_T(\pi / Y_{obs}) = \prod [(1 - \pi)(1 - p) + \pi p]^{y_i^T} [\pi(1 - p) + (1 - \pi)p]^{1-y_i^T}$$

Como consecuencia el estimador de máxima verosimilitud de  $\pi$  es:

$$\hat{\pi}_T = \frac{p - 1 + \lambda}{2p - 1}, \quad (1)$$

donde  $\bar{y}^T = (1/n) \sum y_i^T$ . Se verifica que  $\hat{\pi}_T$  es un estimador insesgado de  $\pi$  y la varianza de  $\hat{\pi}_T$  está dado por:  $Var(\hat{\pi}_T) = \frac{\lambda(1-\lambda)}{n(2p-1)^2}$ ,  $p \neq 0,5$ , donde  $\lambda = (1 - \pi)(1 - p) + \pi p$ .

### Intervalo de Confianza de Wald para $\pi$

Por el Teorema de Límite Central y  $\hat{\pi}_T$  especificado en (1):

$$\frac{\hat{\pi}_T - \pi}{\sqrt{\widehat{Var}(\hat{\pi}_T)}} \sim N(0, 1), \text{ cuando } n \rightarrow \infty.$$

El Intervalo de Confianza, de Wald, para  $\pi$  al  $100(1-\alpha)\%$  de confianza es dado por:

$$\left[ \hat{\pi}_T - Z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\pi}_T)}, \hat{\pi}_T + Z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\pi}_T)} \right]$$

### 3 Resultados y Discusión

El presente estudio fue realizado en la Facultad de Ciencias Matemáticas, de la UNMSM. Se aplicó el modelo de respuesta aleatorizada - el modelo transversal - a los alumnos matriculados el Semestre académico del 2015-I, para investigar la proporción de personas que han copiado en los exámenes de otros estudiantes por lo menos una vez, el porcentaje de personas que han utilizado notas en un examen, el porcentaje de personas que han utilizado drogas para mejorar su rendimiento académico y el porcentaje de personas que han presentado un documento que contienen un párrafo intencionalmente adoptado del trabajo de otra persona, utilizando la teoría del modelo de respuesta no aleatorizada transversal (Tian & Tang 2014). En el diseño muestral, se utilizó el muestreo aleatorio estratificado con afijación proporcional al tamaño de cada estrato (Scheffer & Mendenhall, 2007) considerando como estratos a las Escuelas Académicas Profesionales (E.A.P.) de la FCM. Para el cálculo del tamaño de muestra se consideró un límite para el error de estimación de 5%, con un nivel de confianza del 95% y la información proporcionada por la Dirección Académica de la FCM, de los alumnos matriculados el primer semestre del año académico 2015, el tamaño de muestra considerado fue de 300 alumnos, repartidos en forma proporcional a las cuatro E.A.P. de la FCM. Se consideró a cada E.A.P. como un estrato: Investigación Operativa, Estadística, Computación Científica y Matemática. Los estudiantes fueron seleccionados aleatoriamente hasta completar el tamaño de la muestra igual a 300.

**Las variables elegidas para este estudio fueron las siguientes:**

**1. Personas que han copiado de otros estudiantes por lo menos una vez**

**Preguntas sensibles:**

- “He copiado de otros estudiantes durante un examen (práctica) por lo menos una vez” ( $Y=1$ )
- “Nunca he copiado de otros estudiantes durante un examen (práctica)” ( $Y=0$ )

**Preguntas no sensibles:**

- Nací entre agosto y diciembre ( $W = 1$ )
- No nací entre agosto y diciembre ( $W = 0$ )

**2. Personas que han utilizado notas en un examen**

**Preguntas sensibles:**

- “He utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares)” ( $Y = 1$ )
- “Nunca he utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares)” ( $Y=0$ )

**Preguntas no sensibles:**

- He interrumpido mis estudios universitarios por lo menos una vez ( $W=1$ )
- Nunca interrumpí mis estudios universitarios ( $W= 0$ )

**3. Personas que han utilizado drogas para mejorar su rendimiento académico en un examen o práctica**

**Preguntas sensibles:**

- “ He utilizado drogas para mejorar mi rendimiento académico en un examen o práctica por lo menos una vez” (Y = 1)
- “ Nunca he utilizado drogas para mejorar mi rendimiento académico en un examen o práctica”(Y=0)

**Preguntas no sensibles:**

- Utilizo biblioteca tradicional (W = 1)
- No utilizo biblioteca tradicional (W = 0)

**4. Personas que han presentado un documento que contiene un párrafo intencionalmente adoptado de otra persona****Preguntas sensibles:**

- “He presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona ” (Y = 1)
- “Nunca he he presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona ” (Y = 0)

**Preguntas no sensibles:**

- Me gusta estudiar solo (a) (W=1)
- No me gusta estudiar solo(a) (W=0)

Cada una de las preguntas consideradas anteriormente fueron organizadas en una tabla de contingencia.

**Tamaño de muestra**

De acuerdo a Registros Académicos de la Dirección Académica los Alumnos Matriculados el Semestre 2015 - I en la Facultad de Ciencias Matemáticas de la UNMSM, está dividida en cuatro Escuelas Académico Profesionales(ver Tabla 2), de acuerdo a esta información se decidió considerar a cada Escuela Académico Profesional como un estrato, en total tenemos cuatro estratos o Escuelas. El esquema de muestreo que se utilizó fue el Muestreo Aleatorio Estratificado, con afijación proporcional de acuerdo a la cantidad de alumnos en cada uno de los estratos, donde cada estrato es una Escuela Académico Profesional de la Facultad de Ciencias Matemáticas de la UNMSM.

Cuadro 2: Distribución de los estudiantes de Pre-Grado según Escuela Académico profesional

Semestre 2015-I-FCM-UNMSM		
Escuela Acémico Profesional	Frecuencia	Porcentaje
Matemática	458	30,11 %
Estadística	291	19,13 %
Investigación Operativa	471	30,97 %
Computación Científica	301	19,79 %
Total	1521	100,00 %

Fuente: Dirección Académica - FCM - UNMSM

Se utilizó el Muestreo Aleatorio Estratificado (Scheaffer; Mendenhall; Ott, 2007), con un nivel de confianza del 95 % y un límite para el error de estimación de 5 %, el tamaño de muestra fue de 300 alumnos matriculados el Semestre 2015-I. El modelo de respuesta no aleatorizada transversal requiere un tamaño de muestra más grande que el método convencional. La entrevista se realizó del 25 de mayo al 13 de julio del año 2015. La distribución de la muestra por Escuela Académico Profesional se muestra en la Tabla 3.

Cuadro 3: Distribución de la muestra según Escuela Académico Profesional

Semestre 2015-I - FCM-UNMSM	
Escuela Académico Profesional	Frecuencia
Matemática	90
Estadística	58
Investigación Operativa	93
Computación Científica	59
Total	300

El estimador del modelo transversal de respuesta no aleatorizada para la “Proporción de personas que han copiado de otros estudiantes en un examen o práctica por lo menos una vez”, cuyo estimador fue desarrollado anteriormente (ver ecuación 1). La proporción de los círculos marcados en la muestra es:  $\hat{\lambda} = \bar{Y}_T = \frac{129}{300} = 0,430$ . La probabilidad de que un estudiante nació entre agosto y diciembre es:  $p = P(W = 1) = 0,4167$ , substituyendo en la ecuación (1), tenemos que:  $\hat{\pi}_T = \frac{0,4167-1+0,43}{2*0,4167-1} = 0,92$ . De forma análoga se calculó la proporción de personas que han utilizado notas en los exámenes, donde la proporción de círculos marcados en la muestra es  $\hat{\lambda} = \bar{Y}_T = \frac{143}{300} = 0,477$ . La probabilidad de que un estudiante interrumpió sus estudios es  $p = P(W = 1) = 0,19$ , substituyendo en la ecuación (1), tenemos que:  $\hat{\pi}_T = \frac{0,19-1+0,477}{2*0,19-1} = 0,538$ . De forma análoga se calculó la proporción de personas que han utilizado drogas para mejorar su rendimiento académico. La proporción de los círculos marcados en la muestra es  $\hat{\lambda} = \bar{Y}_T = \frac{120}{300} = 0,40$ . La probabilidad de que un estudiante utilizó biblioteca tradicional es  $p = P(W = 1) = 0,82$ , substituyendo en la ecuación (1), tenemos que:  $\hat{\pi}_T = \frac{0,82-1+0,40}{2*0,82-1} = 0,344$ . La proporción de personas han utilizado drogas para mejorar su rendimiento académico es 0,344. De forma análoga se calculo la proporción de personas que han presentado un trabajo intencionalmente adoptado del trabajo de otra persona, donde la proporción de círculos marcados es:  $\hat{\lambda} = \bar{Y}_T = \frac{122}{300} = 0,407$  y tenemos información de  $p = P(W = 1) = 0,22$  (probabilidad de que a un estudiante le guste estudiar solo), substituyendo en la ecuación (1), tenemos que:  $\hat{\pi}_T = \frac{0,22-1+0,407}{2*0,22-1} = 0,667$ . La proporción de personas han presentado un trabajo intencionalmente adoptado del trabajo de otra persona es 0,667.

Cuadro 4: Resultados del Modelo transversal: estimación y error estándar

Proporción de personas que han	Estimación	Error estándar
copiado en los exámenes de otros estudiantes	0,92	0,179
utilizado notas en los exámenes	0,538	0,047
utilizado drogas para mejorar su rendimiento académico	0,344	0,044
presentado copia adaptada del trabajo de otra persona	0,667	0,051

En la Tabla 4 se observa que utilizando el modelo de respuesta no aleatorizada - modelo transversal - la proporción de personas que han copiado de otros estudiantes por lo menos una vez es 0,92; el 53,8 % de los entrevistados han utilizado notas en los exámenes, el 34,4 % de los entrevistados

han utilizado drogas para mejorar su rendimiento académico y el 66,7 % de los entrevistados han presentado un trabajo intencionalmente adoptado del trabajo de otra persona.

Cuadro 5: Resultados del Método convencional: estimación y error estándar

Proporción de personas que han	Estimación	Error estándar
copiado en los exámenes de otros estudiantes	0,535	0,104
utilizado notas en los exámenes	0,325	0,080
utilizado drogas para mejorar su rendimiento académico	0,015	0,020
presentado copia adaptada del trabajo de otra persona	0,270	0,076

En la Tabla 5 se observa que utilizando el método convencional - la proporción de personas que han copiado de otros estudiantes por lo menos una vez es 0,535; el 32,5 % de los entrevistados han utilizado notas en los exámenes, el 1,5 % de los entrevistados han utilizado drogas para mejorar su rendimiento académico y el 27 % de los entrevistados han presentado un trabajo intencionalmente adoptado del trabajo de otra persona.

## 4 Conclusiones

El 67 % de los alumnos entrevistados pertenecen al género masculino mientras que el 33 % al femenino. La edad promedio de los estudiantes fue de 21,82 años. El modelo de respuesta no aleatorizada - el modelo transversal - permitió implementar una serie de técnicas de muestreo estadístico, que se mostró eficiente en la selección y ejecución de la investigación.

La estimación de la “Proporción de personas que han copiado de otros estudiantes en un examen por lo menos una vez” con el modelo de respuesta no aleatorizada fue de 0,92, mientras que con el método convencional fue de 0,535; la estimación del “Porcentaje de personas que han utilizado notas en los exámenes” fue de 53,80 %, mientras que con el método tradicional fue de 32,5 %; la estimación del “Porcentaje de personas que han utilizado drogas para mejorar su rendimiento académico” fue de 34,4 %, mientras que con el método convencional fue de 1,5 %, y la estimación del “Porcentaje de personas que han presentado un trabajo intencionalmente adoptado del trabajo de otra persona” fue de 66,7 %, mientras que con el método tradicional fue de 27 %.

El modelo de respuesta aleatorizada es más eficiente cuando es utilizada en investigaciones donde el problema es altamente sensible y para el cual se requiere que el tamaño de muestra es mayor que el método tradicional.

De los resultados se concluye que el Modelo Transversal, fue útil y provechosa en una investigación de la realidad de nuestra sociedad, identificar el comportamientos en los estudiantes de la FCM frente a preguntas sensibles, se recomienda seguir aplicando los modelos de respuesta no aleatorizada en investigaciones de nuestra realidad. Las desventajas del MRNA está en el hecho de que los gastos en la capacitación de los entrevistadores (el entrenamiento en la técnica) y el tiempo que requiere la entrevista para explicar la técnica al entrevistado es más alta con respecto al método tradicional. Se recomienda seguir experimentando estos modelos, en muestras más grandes y en temas en donde la pregunta sea efectivamente altamente sensible o muy comprometedor. Estos modelos también se podrían implementar utilizando la inferencia estadística bayesiana.

## Financiamiento

Los autores expresamos nuestro agradecimiento al Consejo Superior de Investigaciones de la Universidad Nacional Mayor de San Marcos por el apoyo financiero para la ejecución del estudio motivo de la presente publicación.

## Referencias Bibliográficas

- [1] Fox, J.A., Tracy, P.E.(1986). *Randomized response: A method for sensitive surveys*. Sage University Paper on Quantitative Applications in the Social Sciences, 58, Beverly Hills: Sage Publications.
- [2] Scheaffer, R., Mendenhall, W. y Ott, L. (2007). *Elementos de muestreo*. Grupo Edit. International Thomson Paraninfo S.A., Madrid, España.
- [3] Solano, O. et al. (2010). *Modelo de resposta aleatorizada:aplicação do modelo de Simmons*. Revista Brasileira de Biometria, São Paulo, Vol 28, N<sup>o</sup>4: 43-51.
- [4] Takahasi, K., Sakasegawa, H. (1977). *A randomized response technique without making use of any randomized device*. Annals of the Institute of Statistical Mathematics. Vol 29:N<sup>o</sup>1:1-8.
- [5] Tan, M., Tian, G.L., Tang, M.L. (2009). *Sample surveys with sensitive questions: a non-randomized response approach*. The American Statistician. Vol 63, N<sup>o</sup>1:9-16.
- [6] Tang, M.L., Tian, G.L., Tang, N.S., Liu, Z.Q. (2009). *A new non-randomized multi-category response model for surveys with a single sensitive question: design and analysis*. Journal of the Korean Statistical Society . Vol 38, N<sup>o</sup>1:339-349.
- [7] Tian, G. L., Tang, M. L., Liu, Z. Q., Tan, M., Tang, N. S. (2011). *Sample size determination for the non-randomized triangular model for sensitive questions in a survey*. Statistical Methods in Medicine Research, 20, 159-173.
- [8] Tian, G.L., Liu, Y. (2014). *Sample size determination for the parallel model in a survey*. Journal of the Korean Statistical Society . Vol 43, 235-249.
- [9] Tian, G.L., Yu, J.W., Tang, M.L., Geng, Z. (2007b). *A new non-randomized model for analyzing sensitive questions with binary outcomes*. Statistics in Medicine. Vol 26, 4238-4252.
- [10] Tian, G.L., Tang, M.L. (2014).. *Incomplete categorical data design: Non-Randomized response techniques for sensitive questions in surveys*. Chapman&Hall/CRC, Boca Raton. New York, United States of America.
- [11] Warner, S.L. (1965). *Randomized response: a survey technique for eliminating evasive answer bias*. Journal of the American Statistical Association. Vol 60, 63-69.
- [12] Yu, J. W., Tian, G. L., Tang, M. L. (2008). *Two new models for survey sampling with sensitive characteristic: design and analysis*. Metrika, Vol 67, 251–263.